

# An Enhanced Feature Selection Method to Predict the Severity in Brain Tumor

Parthiban J, Dr. B. Sathees Kumar

Research Scholar, Bishop Heber College, Trichy, Tamil Nadu, India

**How to cite this paper:** Parthiban J | Dr. B. Sathees Kumar "An Enhanced Feature Selection Method to Predict the Severity in Brain Tumor"

Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1910-1917, <https://doi.org/10.31142/ijtsrd26802>



IJTSRD26802

Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



## ABSTRACT

The level of severity of brain tumor is captured through MRI and then assessed by the physician for their medical interpretation. The facts behind the MRI images are then analyzed by the physician for further medication and follow-up activities. An MRI image composed of large volume of features. It has irrelevant, missing and information which is not certain. In medical data analysis, an MRI image doesn't express facts very clearly to the physician for correct interpretation all the time. It also includes huge amount of redundant information within it. A mathematical model known as rough-set theory has been applied to resolve this problem by eliminating the redundancy in medical image data.

This paper uses a rough set method to find the severity level of the brain tumor of the given MRI image. Rough set feature selection algorithms are applied over the medical image data to select the prominent features. The classification accuracy of the brain tumor can be improved to a better level by using this rough set approach. The prominent features selected through this approach deliver a set of decision rules for the classification task. A search method based on the particle swarm optimization is proposed in this paper for minimizing the attribute set. This approach is compared with previously existing rough set reduction algorithm for finding the accuracy. The results originated from the proposed algorithm is more efficient and can generate decision rules that will better classify the tumor types. The rule-based method provided by the rough-set method delivers classification accuracy in higher level than other smart methods such as fuzzy rule extraction, neural networks, decision trees and Fuzzy Networks like Fuzzy Min-Max Neural Networks.

**KEYWORDS:** Brain tumor; Malignancy Level; Rough sets; Particle swarm optimization; prominent feature selection

## 1. INTRODUCTION

The level of severity in brain tumor decides the nature of the treatment to perform. For lower grade brain tumor, the operation provides the higher survival rate. For other cases, the risk in the surgery and the poor life quality are the factors to consider before the operation. MRI technology helps the physician in finding the real facts about the clinical data before the surgery. The nature of brain tumor is severe, but it happens infrequently. Enough experience is needed to make correct judgments about the medical data. The neuro radiologists have to make qualitative decisions out of their rich experience. The relationship between the severity degrees of the brain tumor can be captured through MRI features and described clearly using rules.

C. Z. Ye et al. [9] focused on several constraints like robustness, missing values, understandability and accuracy. He framed a new rule extraction method based on Fuzzy Min-Max neural networks (FRE-FMMNN) [10, 11]. The proposed methodology uses a multi layer perceptron network trained with a back propagation algorithm, and it makes use of nearest neighborhood method. This method (FRE-FMMNN) provides better predictions in finding the malignancy degree than the other available intelligent methods. C.Z. Ye concentrated only over the classification task. Finding glioma MRI features and the degree of

malignancy is not very easy by having only two rules for predictions.

Medical data composed of irrelevant features, missing values and uncertainties. These are all the factors that complicate the process of finding the malignancy degree in medical data. To find the real facts behind the medical data, it requires how efficiently the analyst handles the incomplete and inconsistent information, and with the various levels of representation of data. The intelligent methods like decision tree, fuzzy theory and neural networks are based on such strong assumptions (Sufficient number of experiments, knowledge about dependencies, probability distributions). Taking decisions from incomplete knowledge or inconsistent information are the tedious task.

The Rough set theory [6] can effectively deals with incompleteness and uncertainty in medical data analysis. It is considered to have the good quality of knowledge to efficiently differentiate the malignancy level in medical data. This algorithm removes the redundant information features and selects a feature subset that has the same discernibility features [6] as the original set. Identification of most prominent subsets is the effective part that improves the quality in decision making process. The rule induction

algorithm [4] followed by rough set methodology generates decision rules, which potentially mine the profound medical knowledge and provides a new medical insight from a new dimension. The rules generated by these methods are very useful for physicians to classify the malignancy level [15] and provides the better understanding about the problem in their hand.

In this paper, we use rough sets to predict the malignancy level of brain tumor. A rough set feature selection methodology is used to select feature subsets that are prominent (The feature set becomes a prominent one because of eliminating the often repeating features and describes the decisions as performed by the original feature set to get improved level of accuracy). The selected prominent features are those that influence the decision concepts and it will be helpful in cause-effect analysis [16]. We introduce a new rough set attribute reduction method, which works based on the PSO (search method). This approach is compared with other rough set approaches. The experimental results show that the reduced subset gained by this approach are more efficient and generates decision rule with better classification performance. This rough set approach can achieve higher classification accuracy than the other intelligent methods[3].

This paper is organized as follows. Section 2 describes the concepts of rough sets. In Section 3 the proposed rough set feature selection algorithm with Particle Swarm Optimization is explained. In Section 4 the reducts using the rough set rule induction algorithm and rule based classification methods are explained. Section 5 describes about the dataset. In Section 6 results are compared with other methods for accuracy analysis. Conclusions are described in Section 7.

## 2. Rough set methodology

Rough set theory [6] efficiently handles the uncertainty and vagueness in data analysis part through the mathematical approach. An object becomes indiscernible because of the limited available information. A rough set theory is characterized by two particle concepts known as lower and upper level approximations generated using the objects indiscernibilities. Here the important problem is the reduction of attributes and the generation of decision rules. Inconsistencies are not aggregated or corrected. Instead of that the lower and upper approximations are computed and rules are induced. The rules [4] are categorized into definite and approximate rules depending on the lower and upper approximations.

### 2.1 Basic theory of rough set concepts

An information system can be defined as  $I = (U, AU\{d\})$ , where  $U$  defined as universe with non-empty set of finite objects.  $A$  is defined as non-empty set of finite condition attributes, and  $d$  is a decision attribute.  $\forall a \in A$  the function defined as  $f_a: U \rightarrow V_a$ , where  $V_a$  is the set of values of  $a$ . If  $p \subseteq A$ , there is an equivalence relation based on association

$$\text{INDE}(P) = \{(x, y) \in U \times U \mid \forall a \in P, f_a(x) = f_a(y)\} \quad (1)$$

The partition of universe  $U$ , produced by  $\text{INDE}(P)$  is denoted as  $U/P$ . If  $(x, y) \in \text{IND}(P)$ , then  $x$  and  $y$  are indiscernible by attributes from  $P$ . The equivalence classes of  $P$ -indiscernibility relation are defined  $[x]_p$ . Let  $X \subseteq U$ , the p-

lower approximation  $\underline{P}X$  and  $\overline{P}X$  upper approximation of set  $X$  can be defined as

$$\underline{P}X = \{x \in U \mid [x]_p \subseteq X\} \quad (2)$$

$$\overline{P}X = \{x \in U \mid [x]_p \cap X \neq \emptyset\} \quad (3)$$

Let  $P, Q \subseteq A$  be equivalence relations over  $U$ , then the positive and the negative regions can be defined as follows

$$\text{POS}_p(Q) = \bigcup_{x \in U/Q} \underline{P}X \quad (4)$$

$$\text{NEG}_p(Q) = U - \bigcup_{x \in U/Q} \overline{P}X \quad (5)$$

$$\text{BND}_p(Q) = \bigcup_{x \in U/Q} \underline{P}X - \bigcup_{x \in U/Q} \overline{P}X \quad (6)$$

The positive region of the partition  $U/Q$  with respect to  $P$ ,  $\text{POS}_p(Q)$ , is the set of all objects that can be classified into blocks of the partition  $U/Q$  by means of  $P$ .  $Q$  depends on  $P$  in a degree  $k (0 \leq k \leq 1)$  denoted  $P \rightarrow_k(Q)$ .

$$K = \gamma_p(Q) = \frac{|\text{POS}_p(Q)|}{|U|} \quad (7)$$

For  $k=1$ ,  $Q$  depends on  $P$ , for  $0 < k < 1$   $Q$  depends partially on  $P$ , For  $k=0$   $P$  doesn't dependent on  $Q$ . When  $P$  is a set of condition attributes and  $Q$  is the decision,  $\gamma_p(Q)$  denotes the quality level of the classification task.

The goal of reducing the attribute[4,5] is to remove redundant attributes, so that the reduced set provides the same quality of classification as the original attribute set. The set of all reducts is denoted as

$$\text{Red} = \{R \subseteq C \mid \gamma_R(D) = \gamma_C(D), \forall B \subseteq R, \gamma_B(D) \neq \gamma_C(D)\} \quad (8)$$

A dataset may contain many attribute reducts, The set of all optimal reduct is as follows

$$\text{Red}_{\min} = \{R \in \text{Red} \mid \forall R' \in \text{Red}, |R| \leq |R'|\} \quad (9)$$

### 2.2 Decision rules

The definition of the decision rule is as follows.

For an expression  $c:(a=v)$  where  $a \in A$  and  $v \in V_a$  is an atomic formula for the decision rule which can be tested for any  $x \in X$ . A basic elementary condition  $c$  can be interpreted as the following equation  $c: U \rightarrow \{\text{true}, \text{false}\}$ . A conjunction  $C$  of  $q$  elementary conditions is defined as  $C = c_1 \wedge c_2 \wedge \dots \wedge c_q$ . The cover of a conjunction  $C$ , denoted as  $[C]$  or  $C|_A$ , is the subset of examples that satisfy the conditions denoted by  $C$ ,  $[c] = \{x \in U: c(x) = \text{true}\}$ , which is mentioned as supporting descriptor. If  $k$  is the concept, the positive cover  $[c]_k^+ = [C] \cap K$  denotes the set of possible examples covered by  $C$ .

A decision rule  $r$  for  $A$  is any expression in the form  $\emptyset \rightarrow (d = v)$  where  $\varphi = c_1 \wedge c_2 \wedge \dots \wedge c_q$  is a conjunction, satisfying  $[\varphi]_k^+ \neq \emptyset$  and  $v \in V_d$ . Where  $V_d$  is the set of values of  $d$ . The attribute value pairs occurring in the left hand side of the rule  $r$  is the condition part,  $\text{Pred}(r)$ , and  $\text{Succ}(r)$  in the right hand side is the decision part.

An object  $u \in U$  is matched by the decision rule,  $\varphi \rightarrow (d = v)$  then we confirm that the rule classifies that the rule classifies  $u$  to decision class  $v$ .  $\text{Match}(r)$  represents the number of objects matched by the decision rule,  $\varphi \rightarrow (d = v)$ , which is equivalent to  $\text{card}(\{u \mid u \in U, \varphi(u) = v\})$ . The support factor of the rule,  $\text{card}(\varphi \mid_A \mid \cap \{d=v\}_A)$  is the number of objects supports the decision rule.

The accuracy and coverage of a decision rule is defined as  $\varphi \rightarrow (d = v)$ .

$$\alpha_{\varphi}(D) = \frac{\text{card}(\{u \mid u \in U, \varphi(u) = v\})}{\text{card}(\{u \mid u \in U\})} \quad (10)$$

$$k_{\varphi}(D) = \frac{\text{card}(\{u \mid u \in U, \varphi(u) = v\})}{\text{card}(\{u \mid u \in U, d=v\})} \quad (11)$$

### 3. Feature selection using rough set [6] with particle swarm optimization technique.

Feature selection through the rough set theory is valuable, that can generate general decision rules [4,5] and provide better classification quality for new samples. However, the problem of finding a minimal reduct is coming under the NP-Hard Category [16]. Heuristic approximation methodology has to be considered for solving the problem. Hu calculates the significance of the attribute using heuristic ideas [14] from discernibility matrices and proposes a heuristic reduction algorithm [DISMAR]. A positive region based method is taken as a significant factor by Hu. A conditional information entropy reduction algorithm [CEAR] was proposed by wang et.al.

We propose a new algorithm to find the minimal rough set reducts using particle swarm optimization on brain tumor data set. The proposed algorithm has been studied and compared with the other rough set reduction algorithms for analyzing accuracy. PSO algorithm performs better for minimal rough set reduction in the experimental results. Kennedy and Eberhart developed the PSO algorithm. It is an evolutionary computation technique. The original idea was to graphically produce the choreography of a collection of birds. PSO was applied effectively to solve the optimization problems. We apply PSO to find out minimal rough set reducts [5].

#### 3.1. PSO algorithm

PSO is initialized with a population of particles. Each and every particle is taken as a point in an S-dimensional space. The best previous position of any particle (Pbest – The position giving the best fitness value) is  $P_i = (p_{i1}, p_{i2}, \dots, p_{is})$ . The index of the global best particle is denoted as 'gbest'. The velocity for any particle  $i$  is  $V_i = (v_{i1}, v_{i2}, \dots, v_{is})$ . According to the following formula the particles are manipulated in the space S.

$$V_{id} = wV_{id} + c_1 \text{rand}() (p_{id} - X_{id}) + c_2 \text{Rand}() (p_{gd} - X_{id}) \quad (12)$$

$$X_{id} = X_{id} + V_{id} \quad (13)$$

In the above equations  $w$  is the inertia weight, suitable selection of inertia weight provides a better balance between global and local exploration and it requires less number of iterations on average to find the optimum value. Time varying inertia will always provides better performance. The acceleration constants are  $c_1$  and  $c_2$  in equation 12, which represent the weighting of stochastic speeding terms that

pull each particle towards Pbest and gbest positions. Low values allow particles to roam far from target regions. High values result in abrupt movement toward target regions.  $\text{Rand}()$  and  $\text{rand}()$  are two random functions in the range [0,1] velocities of each particle on each dimension are limited to a maximum velocity  $V_{\max}$ . If  $V_{\max}$  is smaller, particles may not explore efficiently beyond locally good regions. If  $V_{\max}$  too high particles may fly past the good solutions.

The first part of the eq-12 activates the "flying particles" with able memory capacity and the ability to find a new search space area. The second part is known as the "cognition phase" which represents the private thinking of the particle. "social phase" is the third part, which represents the collaboration among the other particles. Equation-12 is used to update the velocity of the particle [16]. Then the particle flies towards a new place according to the equation-13. The fitness function is used to measure the performance of each particle.

#### The implementation of PSO algorithm is as follows.

1. In S-dimensional problem space, a population of particles with random positions and velocities are initialized. Initialize  $P_i$  with a copy of  $X_i$ , and initialize  $P_g$  with the index of the particle having the best fitness function value among the population.
2. In S-dimensional problem space for each and every particle present, evaluate the desired optimization fitness function in d variables.
3. The particle's fitness evaluation is compared with particle's pbest. If current value of the particle is better than the pbest, then set the pbest value equal to the current value, and the pbest location equal to the current location in d dimensional space.
4. Comparison is done on fitness evaluation [22] with the population's overall previous best value. If current value is better than the gbest, then reset gbest to the current particle's array index and value.
5. According to the formulas 12 and 13 change the velocity and position of the particle in the S-dimensional space.
6. Continue the looping until a criterion is met.
7. Usually a sufficiently good fitness value or a maximum number of iteration for this process.

#### 3.2. The encoding process

The position of the particle [3] is represented as a binary bit of strings with length N. Where N represent the total number of attributes. Every bit is equivalent to an attribute, the value '1' means the corresponding attribute is selected while '0' means the attribute is not selected. Each position is an attribute subset in the problem space.

#### 3.3. Velocity representation

Particle's velocity in the problem space is represented with a positive integer, which varies between 1 and  $V_{\max}$ . It means that at any point in time how many of the particle's [5] bit should be changed to be the same as that of the global best position, which means the velocity of the particle flying toward the best possible position.

For example,  $pgbest = [1011101001]$  is the value and  $X_i = [0100110101]$ . The difference between gbest and the particle's current position is  $pgbest - X_i = [1-1110-11-100]$ . The value '1' means that, compared with the best position, this



bit(feature) should be selected but it is not because it decreases the quality of classification. The value '-1' means that compared with the best position, this bit should not be selected but it is selected. Redundant features will increase the cardinality value of the subset. Both these cases will lead to a lower fitness value. The number of '1's is a and that of '-1' is b. The difference between a and b (a-b) is the distance between the two taken particle's position. The (a-b) value may be positive or negative. Such a variety makes particles possess 'exploration ability' in solution space.

In this example (a-b)=4-3=1, so  $Pg-X_i=1$ .

### 3.4. Strategies to update a particles position

After updating the velocity, a particle's position will be updated with the new velocity value. If the new velocity value is V, the number of different bits between the gbest [21] and the current particle is xg, there is two situations for updating the position.

1.  $V \leq xg$ . If a situation is like this, randomly change V bits of the particle, which are different in form that of gbest. The particle will move toward the global best while keeping its 'searching ability'.
2.  $V > xg$ . If a situation is like this, instead of changing all the different bits to be same as that of gbest, we should randomly change (V-xg) bits outside the different bits between the particle and gbest. After reaching the global best position, it keeps on moving some distance toward other possible directions, which gives it further searching ability.

### 3.5. The velocity limit (maximum velocity, Vmax)

The particle's velocity was initially limited to region [1,N]. It was noticed that in some cases after several iterations, the swarms [3] find a good solution (but not the real optimal one), and in the following generations gbest remains stationary. So, only sub optimal solution is obtained. This is the condition in which the maximum velocity of the particle [19,20] is too high and particle often 'fly past' the optimal solution.

We set  $V_{max}$  as  $(1/3)N$  and limit the range of the velocity in  $[1, (1/3)N]$ , which prevents this from being too large in value. Limiting the maximum velocity, particles cannot fly too far away [3,16] from the optimal solution. After finding a global best position, other particles will adjust velocities and positions and search around the best position. If  $V < 1$ , then  $V=1$ . If  $V > (1/3)N$ ,  $V=(1/3)N$ . PSO [16] can find optimal reducts [5] quickly under such limiting factors.

### 3.6. The fitness function

We apply the fitness function as given below:

$$\text{fitness} = \alpha \gamma R(D) + \beta \frac{|C| - |R|}{|C|} \quad (14)$$

Where  $\gamma R(D)$  decides the classification quality of the condition attribute set R relative to decision D. |R| is the '1' number of a position or the selected feature subset length. |C| is the total number of features,  $\alpha$  and  $\beta$  are two parameters that correspond to the importance of classification quality and subset length, with  $\alpha \in [0,1]$  and  $\beta = 1 - \alpha$ . In our experiment we set  $\alpha = 0.9$  and  $\beta = 0.1$ . The high value of  $\alpha$  assures that the best position [3] is at least a

real rough set reduct. The ultimate goal is to maximize the fitness values.

### 3.7. Setting the parameters

In the given algorithm, the inertia weight decreases along with the iterations according to the equation-15

$$W = W_{\max} - \frac{W_{\max} - W_{\min}}{\text{iter}_{\max}} \text{iter} \quad (15)$$

Where  $W_{\max}$  is the initial value of weighting coefficient,  $W_{\min}$  the final value of the weighting coefficient,  $\text{iter}_{\max}$  the maximum number of assigned iterations [16] or generations, and iter represents the current iteration or generation number.

### 3.8. Measuring the time complexity

Let N be the number of features (representing conditional attributes) and M the total number of objects. The time complexity of POSAR is  $O(NM^2)$ , and that of the reduction [17,18] based on conditional information entropy (CEAR) [6] is  $O(NM^2) + O(N^3)$ , which is composed of computation of the core and non-core attribute reduct. DISMAR has total time complexity  $O((N + \log M)M^2)$ . For PSOREDUC, the complexity of the fitness function is  $O(NM^2)$ , the other impacts are measured by the generation iterations. Time spent on evaluating the particle's position (that is fitness function).

## 4. Rough set rule induction algorithms

### 4.1. Algorithm for induction of minimum set of decision rules.

LEMM algorithm was proposed to extract a minimum set of decision rules. Let K be a non empty lower or upper approximation of a concept, where c is a basic condition, and C denotes the conjunction of such conditions being a candidate for condition part of the decision rule. C[G] represents the currently considered to be added to the conjunction.

Procedure LEMM

(Input: a collection of objects K,

Output: Decision rules R derived out of the algorithm);

begin

G:=K;

R:=  $\emptyset$

while G  $\neq \emptyset$  do

begin

C:=  $\emptyset$

C(G):={ c:[c]  $\cap$  G  $\neq \emptyset$  }

while (C= $\emptyset$ ) or (not [C]  $\subseteq$  K) do

begin

select a pair c  $\in$  C(G) such that |[c]  $\cap$  G| is minimum;

if match occur then select a join up c  $\in$  C(G) with the biggest |[c]|;

if further ties occur then select the last join up from the list

C:= C  $\cup$  {c};

G:= [c]  $\cup$  G;

C(G):={ c:[c]  $\cap$  G  $\neq \emptyset$  }

C(G):= C(G) - C;

end {while}

for each  $c \in C$  do

if  $[C-c] \subseteq K$  then  $C := C - \{c\}$ ;

Create rule  $r$  based on  $C$  and add it to rule set  $R$ ;

$G := K - \bigcup_{r \in R} [R]$ ;

end{ while };

for each  $r \in R$  do

if  $\bigcup_{s \in R-r} [S] = K$  then  $R := R-r$ ;

end {procedure};

#### 4.2 Classification based on Decision rules

The LEMM algorithm is mainly used for classification purpose. The induced set of rules is employed to [12] successfully classify a new object. If new object matches more than one rule, it needs to resolve the conflict between sets of rules classifying tested objects to different decision [16] classes.

In Ref [16], additional coefficients characterizing rules are taken in to account: the strength of matched or partially matched rules (the total number of cases correctly classified by the rule during training phase), the number of non-matched conditions, the rule specificity (i.e. length of the condition parts). All these coefficients are combined together and the strongest decision wins. If no rule is matched, the partially matched rules are taken and the most probable decision is chosen from that.

The global strength defined in Ref. [16] for rule negotiation is a number in the range [0,1] representing the importance of sets of decision rules relative to the considered tested object. Let us assume that  $T = (U, AU\{d\})$  is a given decision

table, Consider  $u_t$  is a test object,  $Rul(X_i)$  is the set of all calculated decision rules [18] for the equation  $T$ , classifying objects to the decision class

$X_i(w_d^j = v_d)$ ,  $MRul(X_i, u_t) \subseteq Rul(X_i)$  matching tested object  $u_t$ . The global strength of decision rule set  $MRul(X_i, u_t)$  is defined as follows

$$Glstrength(X_i, u_t) = \frac{(\text{card}(\bigcup_{r \in MRul(X_i, u_t)} \text{Pred}(r)) \cap \{d = v_d^j\} \cap A)}{\text{card}(\{d = v_d^j\} \cap A)} \quad (16)$$

To classify a new case, rules are first selected matching the new case. The strength of the selected rule sets is calculated for any decision class, and then the decision class with maximal strength is selected, with the new case being classified to this current class. The quality of the complete set of rules on a dataset with size  $n$  is evaluated by the classification accuracy:  $n_c/n$ , where  $n_c$  is the number of examples that have been correctly classified.

#### 5. Brain tumor data set

The brain tumor data set [15] contains 14 condition attributes and one decision attribute, as shown in Table 1. The decision attribute 'clinical grade', is the actual grade of glioma obtained from the surgery. Except the attributes like 'gender', 'age' and 'clinical grade', other attributes are derived from the MRI of the patient and described with uncertainty to various extent levels. The numerical attribute 'age' is discretized into three degrees, 1-30, 31-60, 61-90, represented by 1, 2, 3 respectively.

Table1: Attribute description of brain tumor dataset

No	Label	Attributes	Description
1	a1	Sex	0:female; 1:male
2	a2	Age	1:[<=30]; 2:[31-60]; 3:[>60]
3	a3	Type of the Shape	1:round; 2:ellipse; 3:irregular
4	a4	Nature of Contour	1:clear; 2:partially clear; 3:blur
5	a5	Tumor capsule	1:uninjured; 2:partially injured; 3:absent
6	a6	Edema	0: absent; 1: light; 2: middle; 3: heavy
7	a7	Presence of Mass	0: absent; 1: light; 2: middle; 3: heavy
8	a8	Post-contrast enhancement	-1: unknown; 0: absent; 1:homogenous; 2: heterogeneous
9	a9	Blood supply level	1: normal; 2: middle; 3:affluent
10	a10	Premature death of cells	0: absent; 1: present;
11	a11	Calcification of the tumor	0: absent; 1: present;
12	a12	Nature of Hemorrhage	0: absent; 1: acute; 2: Chronic
13	a13	T1- weighted image's signal intensity	1: hyper intense only; 2: isointense or accompanied by hyper intense; 3: hyperintense or accompanied by isointense
14	a14	T2- weighted image's signal intensity	1: hyper intense only; 2: isointense or accompanied by hyperintense; 3: hyperintense or accompanied by isointense
15	a15	Lesions	-1:unknown 1:present 2:absent
16	a16	Clinical grade	1:low grade 2:high grade

In total, 290 cases of brain glioma [15] are collected and separated into two classes: lower grade and higher grade, in which 174 are of low-grade glioma and 116 are of high-grade. There are 126 cases containing missing values on "post-contrast enhancement". By removing the incomplete 126 cases, the remaining subset of 164 complete cases contains 90 low-grade glioma and 74 high-grade. Investigations are conducted on both the 290 cases and the 164 complete cases without the missing values. The quality of classification for both the 290 and 164 cases data are equal to 1. i.e the positive regions contains all the cases.

## 6. Experiment results

In our experiments, we firstly use the rough set feature selection algorithm to select prominent feature subsets from the brain glioma data [15]. Then, the selected feature subset are applied to generate decision rules to help the neuro radiologists predict the degree of malignancy in brain glioma.

### 6.1 Feature selection and rule set based classification

Matlab 2015a environment is used for implementing the PSOREDUCT [4] algorithm and other types of rough set algorithms. The computer is intel (R) core (TM), 2.20 GHz CPU, 2 GB RAM and the system is windows 7 professional.

We applied the rough set feature selection algorithm to select the prominent features [4] subset from the brain glioma data set. Then the subset of selected features is applied to generate different decision rules to assist the neuro radiologists to predict the malignancy degree [8] in brain tumor.

We use 10-fold cross validation to evaluate the accuracy of classification of the rule set induced from the data set. All cases are re-ordered randomly [16] and then the set of all cases is divided into 10 disjoint subsets approximately in equal size. For each subset, all remaining cases are used for training, i.e., for rule induction, while the other remaining subset is used for testing purpose. Different reordering result in different error rates. So, for each test we perform 10 times 10-fold cross validation and the results are averaged.

The experimental results are listed in Table 2 and Table 3. The parameter settings and their values for PSOREDUCT are in Table 4. We performed experimentation on 290 brain glioma dataset [15] and the 176 case data. For both datasets, decision rules generated from reducts produce a higher classification accuracy than those with the full 15 condition attributes. So, this method of feature selection can improve the accuracy effectively.

The proposed rough set feature selection algorithm is compared against other rough set reduction algorithms. The reducts [4,5] found by our algorithm are more efficient and can generate decision rules with better classification performance. Compared with the other methods provided in the tables, the rough set rule-based classification method achieve higher level in classification accuracy.

**Table2: Results of 290 glioma cases**

Algo. Name	Rough set Reduction algorithms					FRE FMMNN
	All	DISMAR	POSR	CEAR	PSOREDUCT	
Reduct	1-15	2,3,6-9,13,14,15	1-4,6,9,11,13,14,15	2,4,6,7,9,10,12,13,14,15	2,3,5,6,8,9,13,14,15	2,6-9,12,15
Rules	51	54	52	50	51	2
Avg. accuracy (%)	82.80	84.70	83.58	85.49	87.67	83.21
High (%)	96.45	94.48	96.48	100	100	89.39
Low (%)	67.85	60.71	67.85	67.85	64.38	75
STD	7.23	7.11	5.87	6.09	6.63	5.35

**Table3: Experimental results of 164 complete glioma**

Algo. Name	Rough set Reduction algorithms					FRE FMMNN
	All	DISMR	POSR	CEAR	PSOREDUCT	
Reduct	1-15	2,3,6-9,13,14,15	2,3,5-9,11-14,15	1-3,6-9,13,14,15	2,3,5,6,8,9,13,14,15	2,6,8,9,11,13,15
Rules	32	32	26	34	30	2
Avg. accuracy (%)	78.26	82.77	86.77	86.10	87.67	86.37
High (%)	93.63	100	100	100	100	100.00
Low (%)	53.63	53.63	70	80	73.36	73.36
STD	8.78	10.24	8.48	5.48	9.39	8.59

**Table4: PSOREDUCT parameter values**

Parameters	PSO
Population Size	25
Max generation	100
Initial value of c1	2.0
Initial value of c2	2.0
Weight	1.4 - 0.4
Maximal velocity	1-(1/3)N

Features like age, edema, post contrast enhancement, blood supply and signal intensity [15] of the T1-weighted image are the most important factors for predicting the malignancy degree. These results are in accord with the experiences of experts and other researcher's contributions, and are very useful to neuro radiologists.

## 6.2 Decision rules generated using brain glioma data

The results based on the full 290 cases are more useful to neuro radiologists. In Table 6 we present part of the rules extracted from the 290-case brain glioma data. The rules are generated by the rough set rule induction algorithm and include certain rules and possible rules. Rules 1-3 are possible rules and others are certain rules.

**Table 5: Selected feature subset**

Features selected for our approach	
By experiment(experts)	5-10, 12-15
Ye	2,6-9,11-14
PSOREDUCT	2,3,5,6,8,9,13,14,15
Intersection	2,6,8,9,13,15

The three possible rules have the higher accuracy and coverage level. Rule 1, if (post-contrast enhancement is absent) then that denotes the (low-grade brain glioma), covers 55 of 169 low-grade cases and has an accuracy of 98.2 %. Rule 2, If (affluent blood supply) Then (high-grade brain glioma), covers 80 of 111 high grade cases and has an accuracy of 81.6%. Rule 3 shows that hypointense only of signal intensity of the T1- and T2-weighted image always leads to low-grade brain glioma. This rule covers 114 of 169 low grade cases and has an accuracy of 72.61%.

**Table6: Decision rules with measures**

No	Rules	Measures(supp,acc,cov)
1	$(a8=0) \geq (d=1)$	(56,98.2%, 32.5%)
2	$(a9=3) \geq (d=2)$	(81, 81.6%,72.0%)
3	$(a13=1) \& (a14=1) \geq (d=1)$	(115, 72.61%, 67.46%)
4	$(a8=0) \& (a9=1) \geq (d=1)$	(55, 100%, 30.18%)
5	$(a6=1) \& (a9=) \& (a13=1) \geq (d=1)$	(49,100%, 27,81%)
6	$(a2=1,2) \& (a6=0) \& (a9=1) = > (d=1)$	(48, 100%, 24.85%)
7	$(a3=2) \& (a6=0) \& (a13=1) \geq (d=1)$	(23, 100%, 12.43%)
8	$(a3=2) \& (a6=1) \& (a9=1) \geq (d=1)$	(35, 100%, 20.71%)
9	$(a2=1) \& (a5=1) \geq (d=1)$	(19, 100%, 10.65%)
10	$(a2=2) \& (a3=1) \& (a9=1) \& (a13=1) = > (d=1)$	(22, 100%, 11.24%)
11	$(a2=2) \& (a6=2) \& (a8=2) \& (a9=3) \geq (d=2)$	(19, 100%, 10.65%)
12	$(a9=3) \& (a14=3) \geq (d=2)$	(19, 100%, 10.65%)
13	$(a3=3) \& (a9=3) \& (a13=2) \& (a14=1) \geq (d=2)$	(9,100%,4.73%)

Rules starting from 4-13 are considered as certain rules, where rules from 4-10 are for low grade glioma and rules from 11-13 are for high-grade brain glioma. The following conclusions can be drawn from these set of rules.

1. If age is young AND the shape is regular AND edema is absent or light AND post contrast enhancement is absent AND blood supply is normal AND the signal intensity is hypointense for T1 and T2 weighted images then the result is possibly brain glioma with lower grade.
2. If age is old AND the shape is irregular AND edema is heavy AND post contrast enhancement is homogeneous or heterogeneous AND the blood supply is affluent Then the result is most possibly brain glioma with higher grade.

Light edema or the absence of edema often represent the lower grade brain glioma [15], if the edema is heavy in nature, it represent the higher grade brain glioma. If the shape is round or ellipse (Regular in shape) that represent the lower grade case. The normal blood supply and the absence of post contrast enhancement always indicates the lower grade. The affluent supply of blood always indicates a high-grade brain glioma case.

Experimental results are also matches with the medical expert's experiences and other researcher's contributions. These conditional also provides meaningful medical explanations for the physician.

## 7. Conclusions

The rough set theory is applied in this paper to predict the malignancy degree of brain glioma and satisfactory results were obtained. In this paper, attribute reduction using rough set concept is applied with PSO to select the more efficient feature subset. Decision rules were derived from the selected prominent subsets for predicting the degree level of brain glioma. The proposed approach performs well, while comparing with the other rough set methodologies. The reducts found by the proposed approach were more efficient and decision rules were generated based on these reducts provides the better classification accuracy. The degree prediction is based on the features like shape, age, edema, blood supply, post-contrast enhancement and the signal intensity of the T1 and T2 weighted image. The classification accuracy can be improvised by the appropriate feature selection. The rough set based method can achieve good classification accuracy compared with other intelligent techniques.

The decision rule generated by the rough set rule induction approach is useful for both classification and medical knowledge discovery process. These rules effectively expose the interpretable patterns of relations between glioma MRI features and the degree of malignancy. The outcomes of rules are helpful for the medical experts.



## References

- [1] R. B. Bhatt and M. Gopal, "FRID: Fuzzy-Rough Interactive Dichotomizers," IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'04), pp. 1337-1342, 2004.
- [2] Xiangyang wang, Jie yang, Richard jenson, Xiaojun liu, Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma, Computer methods and programs in biomedicine 83 (2006) 147-156
- [3] E. Bonabeau, M. Dorigo, and G. Theraulez, Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press Inc., 1999.
- [4] I. Cloete and J. Van Zyl, "Fuzzy rule induction in a set covering framework," IEEE Transactions on Fuzzy Systems, vol. 14, no. 1, pp. 93-110, 2006.
- [5] W. W. Cohen, "Fast effective rule induction," In Machine Learning: Proceedings of the 12th International Conference, pp. 115-123, 1995.
- [6] C. Cornelis, M. De Cock and A. Radzikowska, "Vaguely Quantified Rough Sets," Proc. 11th Int. Conf. on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007), pp. 87-94, 2007.
- [7] C. Cornelis, G. Hurtado Mart'ın, R. Jensen, and D. Sle,zak, "Feature ´ Selection with Fuzzy Decision Reducts", Proc. 3rd Int. Conf. on Roughset.
- [8] M. Bredel, L.F. Pollack, The p21-Ras signal transduction pathway and growth regulation in human high-grade gliomas, Brain Res. Rev. 29 (1999) 232-249.
- [9] C. Z. Ye, J. Yang, D.Y. Geng, Y. Zhou, N. Y. Chen, Fuzzy rules to predict degree of malignancy in brain glioma, Med. Biol. Comput. Eng. 40 (2) (2002) 145-152.
- [10] P. K. Simpson, Fuzzy Min-Max Neural Networks. Part 2.Clustering, IEEE Trans. Fuzzy Syst. 1 (1993) 32-45.
- [11] J. R. Quinlan, Induction of decision trees, Mach. Learn. 1(1986) 81-106.
- [12] J. M. Zurada, Introduction to Artificial Neural Systems, West Publishing Co., New York, 1992.
- [13] W. Andrew, Statistical Pattern Recognition, Oxford University Press Inc., Oxford, 1999.
- [14] M. A. Lopez-Gonzalez, J. Sotelo, Brain tumors in Mexico: characteristics and prognosis of glioblastoma, Surg. Neurol. 53 (2000) 157-162.
- [15] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, Dordrecht, 1991.
- [16] A. E. Hassanien, Rough set approach for attribute reduction and rule generation: a case of patients with suspected breast cancer, J. Am. Soc. Inform. Sci. Technol. 55 (11). (2004) 954-962.
- [17] Gunther, G., and D. Ivo (2000). Statistical Techniques for Rough Set Data Analysis in Rough Sets: New Developments. Physica-Verlag.
- [18] Kryszkiewicz, M., and H. Rybinski (1996b). Reducing information systems with uncertain real value attributes. In Sixth International Conferences, Informatio Processing and Management of Uncertainty in Knowledge Based Systems, Proceedings (IPMU'96), Vol. II. July 1-5, Grenada. pp. 1165.
- [19] Millonas, M. M. (1994). Swarms, phase transitions, and collective intelligence. In C. G. Langton, Ed., Artijicial Life III. Addison Wesley, Reading, MA.
- [20] Reeves, W. T. (1983). Particle systems - a technique for modeling a class of fuzzy objects. ACM Transactions on Graphics, 2(2):91-108.
- [21] Reynolds, C. W. (1987). Flocks, herds and schools: a distributed behavioral model. Computer Graphics, 2 1 (4):25-34.
- [22] Wilson, E.O. (1975). Sociobiology: The new synthesis. Belknap Press, Cambridge, MA.